DOCUMENT RESUME

ED 337 477 TM 017 291

TITLE Proceedings of the 1980 IPMAAC Conference on Public Personnel Assessment (4th, Boston, Massachusetts, July 6-10, 1980).
INSTITUTION International Personnel Management Association, Washington, DC.
PUB DATE Jul 80
NOTE 43p.
PUB TYPE Collected Works - Conference Proceedings (021)

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS Assessment Centers (Personnel); Evaluation Methods; Job Analysis; *Job Performance; *Occupational Tests; *Personnel Evaluation; Personnel Management; Personnel Selection; Psychometrics; *Public Sector; *Test Use; Workshops
IDENTIFIERS International Personnel Management Association

ABSTRACT

The International Personnel Management Association Assessment Council (IPMAAC) contains over 500 psychometric specialists, personnel psychologists, and personnel staffing specialists dedicated to the improvement of public personnel assessment in such fields as selection and performance evaluation. Author-generated summaries/outlines of papers presented at the IPMAAC's 1980 conference are provided. Four preconference workshops are summarized. The luncheon address is "Government Regulations and Professional Standards: Conflict or Coordination?" by D. J. Schwartz. The presidential address is "Automating the Examination Process in the 80s" by T. S. Darany. Two invited addresses are reviewed: "The Division 14 Principles" by M. Tenopyr; and "Personnel Assessment from the Perspective of Work Adjustment" by R. V. Dawis. The following paper sessions are summarized: "Ratings of Training and Experience"; "Personality Testing"; "Performance Evaluation and Criterion Development"; "Assessment Centers"; "Innovations in Test Validation"; "Physical Testing"; "Issues in Test Fairness and Affirmative Action"; "Alternative Selection Techniques"; and "Various Approaches To Measuring the Cost-Effectiveness of Tests". The following symposia are summarized: "Bias in Content Valid Tests"; "The Southeastern Regional Test Development Center: A New Approach for Consortia"; "Computer-Based Job Analysis--Some Innovative Applications to Personnel Management"; "Issues in the Kirkland Case"; "Clerical Selection and Cooperative Studies"; "Assessing Assessment Research"; "Assessing Writing Skills for Employment Selection"; and "Alternative Selection Procedures for Entry-Level Positions". "Cooperative Job Analysis and Test Development Project for the Class of Conservation Officers" is also summarized as a separate paper. (SLD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

✓ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

MARIANNE ERNESTO

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

IPMA Assessment Council

ED 224427

IPMA ASSESSMENT COUNCIL

PROCEEDINGS OF THE

1980 IPMAAC CONFERENCE

ON

PUBLIC PERSONNEL ASSESSMENT

JULY 6 - 10, 1980

BOSTON, MASSACHUSETTS

TM012291

INTRODUCTION

ABOUT IPMAAC

The International Personnel Management Association Assessment Council (IPMAAC) contains over 500 psychometric specialists, personnel psychologists, and personnel staffing specialists dedicated to the improvement of public personnel assessment in such fields as selection and performance evaluation. The Assessment Council evolved from a Selection Specialists Symposium sponsored in July of 1976. The one hundred and fifty individuals participating in that session supported the establishment of an organization to further intergovernmental communication and cooperation in the area of assessment, with the intention of improving selection practices. In April of 1977, the first IPMAAC annual conference was held in Kansas City, Missouri.

ABOUT THE CONFERENCE

The annual IPMAAC Conference is a major source of communication of ideas among assessment professionals in IPMA. The fourth annual conference, held in Boston, hosted over 100 presenters and approximately 250 attendees.

ABOUT THE PROCEEDINGS

The summaries of presentations in the Proceedings were prepared by members of the IPMAAC Education and Training Committee and others acknowledged below. The purpose of the summaries was to indicate topics addressed and summarize important points.*

Some of the presentations are being prepared also for publication in their entirety as monographs or as articles.

Contributors

Bruce Davey
Bob Marshall
Susan Thrash
David Lookingbill
Jack Lawton
Jan Klein
Debra Koyen
Charles Schultz
Barbara Showers

*NOTE: While every attempt has been made to accurately represent the presentations, persons wishing to quote results should consult directly with the original author.

Presentation Index

		Page
I.	Introduction	i
II.	PRE-CONFERENCE WORKSHOPS	1
	Passing Points	1
	CODAP	2
	How to Develop Job-Related Minimum Qualifications and Ratings of Training and Experience	3
	The Assessment Center Method	5
III.	CONFERENCE SESSIONS	6
	INVITED ADDRESS: "The Division 14 Principles," Mary Tenopyr	6
	Bias in Content Valid Tests (symposium)	8
	The Southeastern Regional Test Development Center: A New Approach for Consortia (symposium)	8
	Cooperative Job Analysis and Test Development Project for the Class of Conservation Officers (paper)	11
	Computer-Based Job Analysis--Some Innovative Applications to Personnel Management (symposium)	13
	Issues in the Kirkland Case (symposium)	15
	Ratings of Training and Experience (paper session)	16
	Personality Testing (paper session)	19
	LUNCHEON ADDRESS: "Government Regulations and Professional Standards: Conflict or Coordination?," Donald J. Schwartz	21
	Performance Evaluation and Criterion Development (paper session)	22
	Assessment Centers (paper session)	24
	Clerical Selection and Cooperative Studies (symposium)	27
	Assessing Assessment Research (symposium)	27
	Innovations in Test Validation (paper session)	29
	Physical Testing (paper session)	30
	Issues in Test Fairness and Affirmative Action (paper session)	32
	Alternative Selection Techniques (paper session)	33
	PRESIDENTIAL ADDRESS: "Automating the Examination Process in the 80s," Theodore S. Darany	35
	INVITED ADDRESS: "Personnel Assessment from the Perspective of Work Adjustment," Rene V. Dawis	36
	Various Approaches to Measuring the Cost-Effectiveness of Tests (paper session)	37
	Assessing Writing Skills for Employment Selection (symposium)	38
	Alternative Selection Procedures for Entry-Level Positions (symposium)	38

WORKSHOP I

Passing Points

Leaders: Barbara Showers, State of Wisconsin
Theodore Darany, San Bernardino County Personnel Division
Glenn McClung, Denver Career Service Authority
Robert Shoop, Missouri Personnel Division
Frank Malinowski, Chicago Department of Personnel
GLAC Representative
Frank Salverline, MAPAC Representative

This workshop provided information and discussion on passing points. The discussions centered around the relevant portions of the Federal Guidelines, a historical perspective, a survey of existing strategies, ranking versus pass-fail, factors to consider when setting passing points and strategies in defending passing points.

Ted Darany, in his opening remarks, described the various approaches to setting passing points including the predictive yield policy, the casual approach, statistical methods and test information models.

Glenn McClung discussed the EEOC guidelines as they relate to the setting of passing points. He also presented the issue of ranking versus pass-fail examinations.

Tom Tyler presented the technique called bootstrapping and Barbara Showers discussed the State of Wisconsin's current procedure for setting passing points. Both procedures provide a statistical method which will give evidence that the selection of any one passing point is appropriate.

Frank Malinowski, City of Chicago, and Dan Salverline, MAPAC representative, shared with the participants descriptions of the various methods used by other jurisdictions.

WORKSHOP II

CODAP
(Comprehensive Occupational Data Analysis Programs)

Leader: Stephen Stroik, Prince George's County Office of Personnel

CODAP is a set of computer programs used to automate, process, organize and report occupational data. The CODAP system was originally developed by the Air Force Human Resources Laboratory in the mid-1960s and has been continuously updated and enhanced since that time. Differing versions are available for use on UNIVAC, IBM, and CDC computer equipment.

CODAP can provide managers with quantified job analysis information relevant to selection, classification, training, evaluation and job design. CODAP quantifies and empirically tests occupational factors over a defined inventory of items (usually task statements), along with background questions, to form clusters or groups, define categories and produce prioritized lists meaningful to managers. Typical examples of reports include composite job descriptions, group difference descriptions, variable summaries or cross-tab reports, and ordered task factor summaries. To aid in producing these products, the system incorporates such data analysis techniques as hierarchical clustering, inter-rater reliability measures, and regression analysis.

The presenters in the day-long workshop provided participants with a basic understanding of the CODAP system and its potential use through lectures providing examples and an extensive question-and-answer period.

WORKSHOP III

How to Develop Job-Related Minimum Qualifications and Ratings of Training and Experience

Leaders: Nancy Abrams, U.S. Office of Personnel Management, New York Region
Louis M. Laguardia, U.S. Office of Personnel Management, New York Region
Leroy Sheibley, Pennsylvania State Civil Service Commission

The purpose was to expose participants to several methods and theories of minimum qualification and T&E development and to give participants an opportunity to construct a T&E with guidance and critique from the workshop leaders.

The presentation on minimum qualifications began by identifying some of the traditional uses and abuses of minimum qualification statements. During the presentation, the following definitions of an MQ were offered:

1) An MQ is any specific characteristic or attribute which job applicants are required to have in order to be allowed to compete further in the selection process.

2) MQs should reasonably sort out those applicants who have a reasonable chance of performing or learning to perform the job from those who have little likelihood of success.

3) "Minimum qualifications required of job applicants should identify those elements of training, experience, special skills, or other personal attributes which are <u>essential</u> to performance of the job, and which must therefore be possessed by an applicant before he can reasonably be expected to perform a job satisfactorily." Tennessee State Department of Personnel. <u>Technical Standards for Determining Minimum Qualifications and Examinations Weights</u>.

It was also recommended that a "good" MQ should possess the following characteristics:

Objectivity--no subjective standards;

Validity--clearly linked to specific work performed or specific job requirements;

Reliability--judgments should be consistent;

Acceptability--the logic and/or validity evidence should be in a form which can be understood by unions, operating agencies, applicants, etc.;

Have some cost/benefit value;

Be verifiable; and

Be developed in such a way that all reasonable options are considered.

It was also stressed that, like any other selection device, minimum qualifications need to be based on the results of a careful job analysis.

After a short presentation on the background of training and experience (T&E) ratings, the following standards for T&E development were offered:

1) Based on job analysis;

2) Evaluate most important entry-level work behaviors or job requirements that differentiate superior workers from adequate workers;

3) Use supplemental form to collect information from applicants;

4) Use structured and well-defined rating procedure (e.g., behaviorally defined scales) for personnel selection specialist or SMEs to score supplements;

5) Be reliable, valid, and have no adverse impact.

Working systems for T&E ratings in the states of Pennsylvania and Wisconsin were reviewed as were a task inventory checklist and a questionnaire developed following the methodology outlined in the B.R.E. Exam Preparation Manual. Both of the latter two systems are used to some extent in the Federal system.

In the afternoon, three groups were formed and a T&E rating guideline/questionnaire was developed by each group for the class Employment Service Manager. The afternoon concluded with a critique by the workshop leaders of each group's product.

WORKSHOP IV

The Assessment Center Method

Leader: Dale R. Baker, U.S. Office of Personnel Management

This workshop presented information and participation exercises in developing and operating an assessment center. The assessment center method is highly job-related, has low adverse impact, and is especially useful as a selection device in situations where the employer cannot use a person's past performance as a worker to predict supervisory and management skills. Some key points in developing and operating a center are:

1) Determine objectives of the programs. Management support is very important.

2) Define critical skills. What is "leadership" in a particular job?

3) Develop measurement exercises. Tailor them to the organization and also to the situations, e.g., oral versus in-basket.

4) Select assessors, usually three observers to six candidates. Demand the best managers for two weeks. Balance according to workforce, candidate pool, and easy and tough raters.

5) Train assessors. Training is especially essential in assessment centers because standards of performance are not precisely documented. Team assessment helps, too, to reduce bias.

6) Select an assessment center site away from the raters workplace to avoid conflicting commitments.

7) Conduct assessment center using a standardized approach. Tell candidates in advance what is going to be assessed. Results tend to be constant for repeaters so advance knowledge does not give advantage.

8) Write reports. Allow one and a half hours per candidate.

9) Use reports. One to two hours of feedback to a candidate in a controlled situation is essential.

10) Evaluate the assessment center.

INVITED ADDRESS

The Division 14 Principles

Mary L. Tenopyr, AT&T

In a presentation reminiscent of Lee Cronbach's 1979 IPMAAC keynote address, Dr. Tenopyr criticized the "trichotomy of the three Cs--criterion-related, content and construct validity" and offered a glimpse of a newer, more flexible conceptualization of validity which influenced the thinking of the authors of the newly completed APA Division 14 Principles.

Dr. Tenopyr, tracing the history of the "three Cs," noted that the APA's original 1954 testing standards introduced them as types of validity. This suggested an independence of the three validation approaches which does not really exist; in fact, they overlap considerably, and aspects of all three might be demonstrated for the same testing instrument.

Dr. Tenopyr expressed the opinion that even though later versions of the APA test standards (1966, 1974) were restructured to make more clear the interrelatedness of the three Cs and to refer to them as aspects rather than types, this model is too rigid for today's testing environment. She feels the 1978 Uniform Selection Guidelines still treat criterion-related, content and construct validity as three distinct strategies, each with a rigid procedural checklist to follow in order to satisfy Federal/legal requirements. She pointed out that while the checklist-like aspect of the Guidelines is helpful for judges and lawyers, there is too much rigidity in this approach. A test can have an array of sound validity evidence of several "types," yet not completely satisfy the Uniform Guidelines checklist for any single type. Additionally, the section of the Guidelines on construct validity is inadequate and "bears no resemblance to professional writings in the area."

As a more flexible alternative to the three Cs, Dr. Tenopyr discussed the concept of internal/external validity. Internal validity refers to the correspondence which can be inferred between test performance and performance in a defined area. Such inferences are heavily dependent on the investigative and developmental efforts involved in constructing the test. The closer the correspondence between the performance(s) elicited by the test and the performance the test is designed to match, the greater the internal validity. This is, of course, analogous to content validity.

External validity refers to evidence of the generalizability of test scores to other outside criteria, including diverse criteria which are dissimilar on the surface to the test items or tasks. This type of validity, which overlaps both construct and criterion-related validity, is described by Dr. Tenopyr as more robust; whereas a test with high internal validity is more situation-sound, a test with high external validity is more durable and versatile.

Dr. Tenopyr speculated that in a world of rapidly-changing jobs, the government may be doing itself a disservice by concerning itself with internal validity, which is narrow and not as generalizable.

SYMPOSIA

Bias in Content Valid Tests

Moderator: Donald J. Schwartz, Equal Employment Opportunity Commission

Participants: Nancy Abrams, U.S. Office of Personnel Management, New York Region
John Veres, Auburn University at Montgomery
Theresa Holland, U.S. Department of Justice

Discussant: Donald J. Schwartz, Equal Employment Opportunity Commission

Nancy Abrams presented her study of the effects of content bias on predictive validity. She created two content-biased tests of general knowledge of famous people and administered them to black and white students. One tested information about famous blacks, the other about famous whites. The mean test scores of whites were greater than blacks on the white-oriented test, and the mean test scores of blacks were greater than whites on the black-oriented test. Both tests showed low predictive validity of grade point average.

John Veres presented a study of differences in important ratings of tasks between blacks and whites. He found that there were significant differences in the perception of "adequate" work on several jobs. These differences were found to be large enough to affect test content. Paradoxically, however, he found that if the content emphasis of black raters was followed, the result would be to increase the adverse impact of the test.

Theresa Holland advised that the courts look to the written word and a common-sense approach when evaluating validity. Content validity does not look as objective as other types; also, it is a further disadvantage that the testing profession has failed to develop a consensus on validation. Since all sorts of conflicting testimony is available on validation, the courts are looking to evidence of bias. If there is evidence of bias in either the test development or results, it is an automatic red flag to the court.

The Southeastern Regional Test Development Center: A New Approach for Consortia

Moderator: James C. Johnson, Tennessee State Department of Personnel

Participants: *James C. Johnson, Tennessee State Department of Personnel
*Doris Maye, Georgia Merit System
*Robert E. Lockwood, Southeastern Regional Test Development Center

Discussant: Charles F. Sproule, Pennsylvania State Civil Service Commission

*Paper distributed.

The Southeastern Regional Test Development Center was established in 1977 with funding from the U.S. Office of Personnel Management through an IPA Discretionary Grant to the State of Georgia on behalf of Florida, South Carolina, North Carolina, Alabama, Tennessee, Kentucky, Georgia, and Mississippi to provide multijurisdictional test development activities. Technical assistance is provided by the Southern Regional Office of the Educational Testing Service, a contracted-for, private consultant.

The Past

Mr. Johnson presented the background regarding the planning and establishment of the consortium. In part, the Center replaced an earlier consortium, SPARK, which was a state police effort primarily, that become too costly to operate.

Most of the other consortia across the country examined at the time were supported by IPA and under the direction of OPM. Recent history had shown that cooperative efforts were judged to be attractive at the outset because of the efficiency, speed, and value that such unions promised. However, in reality, results often proved to be less than satisfactory. This was due to a combination of technical and management problems such as:

1. Deficiencies in job analysis methods,
2. Differences in applicant populations,
3. Differences in operating conditions,
4. Differences in theoretical perspectives,
5. Differences in test development priorities,
6. Inconsistencies with regard to administrative needs, and
7. Difficulties in coordination among jurisdictions.

The impetus for establishing the Center came from two sources: first, from the legal and legislative pressures placed on states to develop valid examinations, and, second, from the adoption by the State of Tennessee of a strategy to develop new exams using a computer-based system. Mr. Johnson then went on to explain the reasons for the Tennessee system, its key facets, and the regional applications of the system.

Important to this system is the linking together of prepackaged task inventories and item banks based on a transportability model which centers on job and test components. In application, tests could be molded to specific jobs and applicant populations as opposed to doing a common job analysis to develop a common test.

The Present

Ms. Maye addressed the general operation of the Center. The organization of the Center includes an Executive Board (the eight State Personnel Directors), a Technical Committee comprised of Test Development Supervisors, and the Center Staff at ETS. The Executive Board provides policy direction, establishes work priorities, and monitors results; the Technical Committee advises the Board and serves as a technical liaision with the Center Staff who conduct the Center's activities.

The issues facing the Center are common but their solutions, from a Regional Test Development Center perspective, are necessarily unique. Those issues included developing written tests for entry classes requiring either specific knowledges or skills or merely the ability to learn to perform the job under job-specific conditions, analyzing item/test data and equating all examinations within a job family, and establishing cutting scores for selection examinations. Novel approaches for each issue were briefly noted and can be examined in Ms. Maye's paper.

To summarize the present situation, the Regional Center concept versus reality was discussed. Operationally, the Center has experienced difficulty in getting the states to work together, to agree on priorities, or to commit resources--staff time and money. Also, the Center Staff has not always been able to respond quickly and their focus (and, thus, the Center's) has been primarily on written examinations, which suggests a reassessment of the Center's purposes.

The Future

Mr. Lockwood introduced the Center's goals by first discussing the "reality factors" relating to the Center's operations and its long-term economic viability. Those factors, cited by Ms. Maye, prompted a reassessment of the Center's purposes, procedures, and priorities. Rather than trying to establish parity among the participating states in technical expertise (i.e. job analysis, item writing, etc.), now the Center will provide, to both member jurisdictions and other "user" jurisdictions, services in the form of test/task banking, test development, and technical assistance in selection beyond written tests.

The expansion to new jurisdictions poses funding match, test security, and user "benefit" issues that must be resolved if this approach is to prove viable. Specifics discussed included developing a data bank, mechanisms to demonstrate transportability, common task statements, standard rating scales, flexible linkages (between task statements and test items), sampling strategy, and item writer training.

Editor's Note: The Executive Board met after the IPMAAC conference to reorganize the Center. One of the major changes to be implemented this fall is the assumption of the private consultant's role by the Technical Committee and the Executive Board after the present contract with ETS expires. The future emphasis will be on interchange of information among consortium members and between the Center and other consortia.

PAPER

Cooperative Job Analysis and Test Development Project for the Class of Conservation Officers

Authors: Bruce Davey, Connecticut State Personnel Division
Phillip Zunder, Vermont Department of Personnel

At the September 1975 meeting of NEPPC, the examination group began work on a cooperative test validation project for the class of Conservation Officer (also called Game Warden in some states). Three major project objectives were identified:

(1) To pool experience and expertise across six states to produce a sound job analysis/test validation methodology.

(2) To provide a vehicle for training and "firing line" experience in test validation in a region where test validation was at an early stage of development.

(3) To explore the advantages and disadvantages of interjurisdictional cooperative projects.

From among 100 benchmark classes provided by the NEPPC classification group, each state in the examination group selected ten classes for which they thought validation was desirable. "Conservation Officer" appeared on every list and was the most consistently high. In this way, a class of common concern was selected. The major phases of the project were as follows:

1. A job analysis phase which included: task analysis and KSAP identification in each state, followed by a joint meeting to produce one comprehensive, common set of task statements and KSAP statements; development of a questionnaire format by a subcommittee after full-group discussion of its desirable elements; and mailing of the questionnaire to more than 200 incumbents and supervisors. This phase showed that there is great similarity in the tasks performed and the knowledges, skills, abilities, and personal characteristics required from state to state, but that some significant differences did exist.

2. A test development phase in which test material was requested of more than fifty jurisdictions, of which fifteen responded, providing more than 3,000 items (before removal of duplicates). These were reviewed, edited, tentatively categorized into the KSAP areas identified in the job analysis, and a total of 371 were selected for purposes of thorough review by subject matter experts.

3. A test validation phase in which conservation supervisors in each state provided further verification of the initial job analysis data; linked the KSAPs to the tasks for which they were required; and reviewed and rated each question for its relevance to the subject area(s), importance to the job, difficulty, and soundness (clarity, accuracy).

The concrete products of the study are:

(1) Training of NEPPC members in job analysis and test development and validation. This is a factor in the substantial increase in validation activities throughout the New England region.

(2) A job analysis report for each state including information on the importance of 48 tasks and 54 KSAPs, and a tie-in between examination areas and tasks performed.

(3) A pool of more than 300 thoroughly reviewed test questions and several thousand others which can potentially be reviewed in the future.

(4) Evidence of the equivalence of some aspects of the Conservation Officer job in the New England region so that other selection devices (physical performance test; oral examination; experience/training evaluation) can be validated in one region and transported to others.

(5) Reusable products (questionnaire format, question rating form) and techniques (data analysis approaches used) for each jurisdiction involved.

(6) Awareness of the strengths and limitations of interjurisdictional validation projects.

SYMPOSIA

Computer-Based Job Analysis--
Some Innovative Applications to Personnel Management

Moderator: Albert P. Maslow, Educational Testing Service

Participants: *Marvin H. Trattner, U.S. Office of Personnel Management
*Robert E. Lockwood, Educational Testing Service
Phyllis Mayes, South Carolina State Department of Personnel
Michael Frank, National League of Cities

Discussant: W. Phalen, United States Air Force

Introduction

Dr. Maslow noted in a preface to the participants' presentations that several applications of computer-based job analysis, based on the CODAP system, have been reported on previously. In each case, they were constrained by the kinds of data gathered and by the existing programs for analysis and display of job information. This symposium highlighted four current programs which represented efforts to extend current concepts and technology to important personnel management problems which depend largely on basic job analysis information.

Applications of CODAP in the Federal Service

Mr. Trattner offered a summary of the current uses of CODAP in the federal government by agency, occupation, and primary product (i.e. selection procedure, training program, etc.). According to a recent survey, one half of the users indicated interest in continuing CODAP use. Those uses include designing training courses (primary military application), validating training, clustering to identify job subspecialities, and assisting in employee career management.

Opposition to CODAP, however, is strong when it provides data on a function that is already being formed. Consequently, the applications are usually for new programs or positions. Mr. Trattner hinted too that many applications were inefficient in terms of their cost--and that a central agency to provide technical/liaison services to CODAP users is needed.

Applications of Computer-Based Job Analysis--Test/Job Linkage

Mr. Lockwood described the Southeastern Regional Test Development Center's attempt to develop a single task inventory from 31 job titles for constructing a core examination for all 31 positions. Six three-point rating scales were developed and used in a multiple-matrix approach to gather data. Importance, need for entry, and time spent scales were placed in one grouping; complexity, degree of personal interaction, and nature of supervision comprised the other grouping.

In each case, 300-400 tasks were rated in a procedure based on a filtering, sorting system less complex than CODAP. Following the evaluation of tasks, the SPSS package was used to develop a preliminary set of test specifications.

In the final product, the personal interaction and nature of supervision scales were dropped.

Mr. Phalen commented that the U.S. Air Force used seven-point scales and avoided combining scales. The time-on-task is obviously relative and that only the time spent scale rather than the importance (do/don't do) scale is used.

The Process of a Multi-Purpose Job Analysis System

Ms. Mayes introduced a common problem: how to do only one data collection for both the classification and testing of new positions. South Carolina's solution was to create an Occupational Standards Unit with responsibility for creating and implementing a job analysis system.

A comprehensive task list was developed from interviews with job incumbents, analyst evaluations, and other jurisdictions' contributions. Tasks were rated on three scales by a panel of randomly selected employees with adequate race and sex representation. The scales included importance (performance differentiation), when performed, and time spent relative to other tasks' times. The results indicated that 80 percent of the jobs surveyed fall neatly into clusters; the remaining 20 percent require additional information.

The system allows both classification and testing work to proceed simultaneously. Currently, the Unit is involved with piloting the system on 10,000 clerical positions in which a sample of 2,000 employees was recently administered the task inventory in one 24-hour period using closed-circuit television.

Technology Transfer (TT) Problems

Mr. Frank defines TT as the process by which an organization adopts innovations--and he currently is involved in an innovative personnel grant for the National League of Cities. This grant involves developing computer-assisted human resource management systems in the areas of classification, performance evaluation, productivity measurement, selection, and training.

Initial research findings indicate that computer applications are limited mainly to record processing and the CODAP is the only successful technology transfer at present. Problems encountered thus far include the modification of technology to novel settings, the disappearance of technology when it is not subsidized with grants and the like, and the fact that policy leaders are often ill-equipped to understand and deal with new technologies (the "protection of one's own 'turf' syndrome").

Summarizing Question

Dr. Maslow noted that the computer indeed helps to deal with specificity. However, is not generalization more important?

Issues in the Kirkland Case

Moderator: Grace H. Wright, New York State Department of Civil Service

Participants: Ken Siegel, New York State Department of Civil Service
Sam Taylor, New York State Department of Civil Service
Judith Gordon, New York State Department of Law

The technical issues in the Kirkland case were discussed in this symposium. Since there was an appeal pending as of the date of the conference, the discussants were unable to present some of the material they had originally planned to present.

The basic issue of the Kirkland case was whether or not the test for Correction Sergeants was valid. The court had decided that an examination for Correction Sergeants given in October 1972 was not substantiated by a credible job analysis. New York proceeded to develop a new examination with an appropriate level of documentation. Their analysis of the 1974 order included the following statement:

> In general, the decision implies that documented, timely knowledge of the tasks performed in a job, how the tasks are performed, and the performance standard, are the base upon which any selection program rests. No other considerations can save a challenged examination, failing adequate documentation of this simple requirement. In addition, every examination element must bear scrutiny as clearly and explicitly related to the job requirements, and the logic of that relation must be demonstrable. This is what "professionally competent" means. The emphasis is now on selection as a technical discipline, subject to technical standards of proof.

The discussion of this symposium centered around the methodology employed by New York in the development of the new examination and the documentation of the validity of the examination. The final order in 1980 was, as excerpted:

> 2. Examination No. 36-435 is approved as criterion valid under the Uniform Guidelines on Employee Selection Procedures as promulgated in 43 (166) Fed. Reg. 38290 et seg. (August 25, 1978);
>
> 3. Examination No. 36-435 is a competitive examination under the New York Constitution Article V, S 6, and Civil Service Law SS 50 (1), (2), 52.
>
> 4. The scoring of candidate performance on Examination No. 36-435 shall incorporate the concept of fairness, or differential validity, set forth in the Uniform Guidelines S 14 (b), (8) (a), (d), 43 (166) Fed. Reg. 38301 (August 25, 1978) and in the Regulations of the Department of Civil Service S 67.1(g), and defendants shall add 250 points to the composite scores (weighted raw score as set forth in Table 5 of the report) for black and Hispanic candidates on Examination No. 36-435 to implement this concept.

PAPER SESSIONS

Ratings of Training and Experience

Chair: Marianne Bays, U.S. Office of Personnel Management, Eastern Region

Presentations: Rating Education, Training, and Experience in the Public Sector
Cindy Lorentson Cook, Indiana State Personnel Division

When is a T and E Rating Valid?
James C. Johnson, Tennessee State Department of Personnel (co-authors: William L. Guffey and Robert A. Perry, Tennessee State Department of Personnel)

Self-Report Questionnaire for Selecting Typists: A Content Validity Approach
Thomas J. Lyons, U.S. Office of Personnel Management

Construction of Reference Checklists from Job Element Study Results
Barry Knake, U.S. Office of Personnel Management, Seattle Region (not presented)

Discussant: Garth Wall, U.S. Army Civilian Personnel Center

Rating Education, Training, and Experience in the Public Sector

This presentation reviewed the results of a questionnaire survey of E&T usage in state, city, and county jurisdictions.

Although little has been written on the subject of E&Ts, use of some type of E&T process appears to be widespread. E&T rating schemes range from very informal and unsystematic, to highly structured and very systematic.

E&T evaluations are based largely on the assumptions that past behavior is the best predictor of future behavior and that accrual of experience in an occupation increases commitment to that occupation.

The survey (sent to TRADE project members) showed that E&T ratings were used more than any other selection device except written tests. They are used both promotionally and open competively at various levels of work and for nearly all occupational areas except public safety-fire/police, clerical, and correctional areas. Economy and speed appeared to be major considerations in choosing E&Ts over other methods of selection.

An E&T rating is generally based around four factors: quantity, quality, relevance, and recency. Recency, however, seems to be approached cautiously in relation to both education and experience.

Although the traditional approach is the most popular, other popular techniques include qualified/not qualified, task element, achievement history questionnaire, and the basic grouping procedure.

When Is a T and E Rating Valid?

Faced with the need to overhaul the current T and E rating procedure, Tennessee initiated a project to develop detailed models to follow and to study the psychometric properties of the T and E rating schemes derived from the models.

The project was based on a concept of validity which defines content validity as the extent to which an examination samples behaviors which are predictive of job performance. Thus, it is the method used to develop an examination, not the examination itself, which must be subjected to empirical validation research.

Ideally, only those indicants would be used which in fact distinguish between more competent and less competent applicants. In practice, we must rely on expert judgment and prior experience with employees having these indicants to tell us whether or not these relationships exist. For instance, if it is not reasonable/rational to assume that two years of "X" training or experience will result in less competence than three years of "X" experience or training, then different examination scores should not result. Therefore, it seems reasonable that the types of indicants found in traditional T&Es may more appropriately be used as minimum qualifications than as a means for rank ordering applicants.

There are, however, some conditions under which it may be acceptable to use a traditional T&E rating scheme. The following are suggested:

1. When the variation in competencies of applicants is very large.
2. When the indicants of training or of work experience are clearly and exclusively linked to job performance potential.
3. When a selection ratio (the ratio of openings to applicants) is extremely large or extremely small.

Finally, three models were presented and discussed. These were a task based model, a KSAP based model, and a behavioral achievement based (behavioral consistency) model.

Task based model. There are several advantages with this model. Tasks provide a more direct description of the job and provide a closer link from a common-sense perspective. The final device is both easier to develop and easier for applicants to understand and complete. An inventory version is easy to score and could be readily automated. Finally, individual indicants could be studied in much the same way as items in a written, multiple choice examination. The task model appears to be most appropriate for jobs that are relatively structured and in which variation in performance is not as important as whether the employee can or cannot perform the task.

KSAP based model. The job element approach is widely used relative to this model. Although this method is sometimes difficult to use and self-ratings have in some cases produced marginal results, Primoff and others report substantial support for using this model for trades and blue-collar work. The KSAP model is more appropriate when competence is acquired in a variety of ways and when knowledges or willingnesses clearly differentiate successful from less successful performance.

Behavioral model. The behavioral model appears to be the most strongly supported model, both theoretically and empirically.

Data were presented from a study for the class Senior Eligibility Worker. In this study, T&E schemes were constructed using all three models and the results were compared.

Self-Report Questionnaire for Selecting Typists: A Content Validity Approach

This study was undertaken to develop an alternative procedure for selecting clerk-typists and clerk-stenographers for Federal employment. The objective of the study was to develop a questionnaire, containing constructed response items for screening applicants, that was conducive to op-scan processing.

Job analysis information used to document the job relatedness of the questionnaire was generated by a CODAP study (Gandy and Maier, 1979) and job element studies of clerk-typist and clerk-stenographer positions in the Federal system.

An item pool was constructed and reviewed by fourteen subject matter experts (SME). On the basis of these ratings, a questionnaire was constructed and tried out on a sample of 174 applicants. On the basis of the tryout sample, a questionnaire and scoring procedure were prepared for a pilot study. The pilot study was conducted in two OPM examining offices.

Initial results seem to indicate that the questionnaire scoring procedures do not have adverse impact on black minority group members.

Personality Testing

Chair: Sue McDonald, State of Florida, Department of Health and Rehabilitation Services

Presentations: Relative Utility of Aptitude and Personality Trait Information in Police Personnel Selection
Barry Morstain, University of Delaware

Psychological Screening of Correction Officers
Robin Inwald, New York City Department of Corrections

Relative Utility of Aptitude and Personality Trait Information in Police Personnel Selection

This paper reports a comparison of aptitude and personality tests for predicting performance ratings of police officers. The findings are summarized in the table below. They show that aptitude and personality tests were equally valuable for predicting performance ratings.

Aptitude Tests	Aptitude Battery	Personality Battery	Combined Battery	Rank
Problem Sensitivity	.22		.22	1
Verbal Comprehension	.26			
Spatial Scanning	.28		.33	4
(All other subtests)	.30			
Personality Tests				
Personal Integration		.20	.27	2
Religious Orientation (Liberal high, fundamental low)		.26	.31	3
Altruism		.28	.34	5
Social Extroversion		.30	---	
Autonomy		---	.36	6
(All other subtests)		.33	.41	

The results showed that three subtests from either battery yielded a multiple correlation of .28, significant at the .05 level. Combining the three best tests from each battery yielded a multiple correlation of .34. The investigator played down the relationship as accounting for only 12 percent of the variance, although Brogden showed in 1946 that a validity coefficient this size would increase productivity substantially, specifically one-third

times the standard test score of those selected times the standard deviation of the measure of productivity ($Y' = r_{xy}z_x$). The predicted performance equals the validity coefficient times the average test score of those selected.

The non-cross-validated multiple correlation using all variables from both batteries was .41.

Psychological Screening of Correction Officers

Robin Inwald presented the results of two studies of psychological screening of corrections officers, as described in the abstracts below.

Pre-Employment Psychological Evaluation as a Predictor of Correction Officer Job Performance

In investigating the predictive validity of psychological screening in law enforcement, a study of the performance of 650 Correction Officers was conducted. Prior to recruitment, these officers were given an MMPI, a biographical questionnaire, an interview by an experienced officer, and, in some cases, an interview by a counseling psychologist. Of the 650, 448 were rated "acceptable"; 164, "marginal"; and 38, "not acceptable." Data were then obtained for the first ten months of service regarding disciplinary actions, corrective interviews, absenteeism and lateness. When the three groups were analyzed and compared for negative job behaviors, it was found that those rated "not acceptable" had significantly more corrective interviews, absence, and lateness than did those rated "marginal" or "acceptable." The usefulness of psychological screening of Correction Officers is discussed.

Effect of Broken Home Background on Race Differences in Psychological Testing

In continuing the investigation of racial differences in personality profiles, 395 male Correction Officer candidates (141 white, 173 black, and 81 Hispanic) completed the MMPI and IPI (Inwald Personality Inventory). Significant differences between races were found on the L, F, K, Hs, Pt, and Sc scales of the MMPI, and AS, PH, AN, IC, ID, and UT scales of the IPI. When scores were analyzed for the subgroup of 258 candidates who reported being raised in two-parent homes, differences remained on only two scales. Hispanics scored higher than the other two groups on the L scale, and blacks scored higher on the UT scale. It is suggested that further consideration be given to family background characteristics when exploring racial differences in personality testing.

LUNCHEON ADDRESS

Government Regulations and Professional Standards: Conflict or Coordination?

Donald J. Schwartz
Equal Employment Opportunity Commission

Dr. Schwartz discussed his seven-year involvement with the development of the Uniform Guidelines on Employee Selection Procedures and the changes and benefits that resulted from the intense interchange between the government and the psychological profession in those seven years.

One outcome of those seven years was the statement of technical standards in the Guidelines which he feels represents a broader consensus of professional opinion than any other document. He stated, "I know of no individual or organization who was denied a request to present their views at a meeting of the principals or the staff committee."

He expressed concern at the dissent of Division 14 of APA in 1979 regarding the Guidelines and called into question the representativeness of the Division 14 Principles and the division's apparent changes in position on validation methods over the years. He suggested that the 1980 Principles not be published until there were open hearings in which all interested parties could participate, or at least until the advisory committee had met in person and resolved all issues.

He suggested that the psychological profession not attempt to use "professional standards" to undercut the government's effort to enforce the law in combating employment discrimination.

In closing, he urged the profession to consider the reasons for the Civil Rights Act and to turn research efforts toward helping to solve the problem of discrimination, not ignore it.

PAPER SESSIONS

Performance Evaluation and Criterion Development

Chair: James C. Johnson, Tennessee State Department of Personnel

Presentations: Quasi-Experimental Pilot Study of a PARTICIPATIVE Performance Appraisal System in Six Washington State Agencies: Effects Upon Job Satisfaction, Agency Climate, and Work Values
Nicholas P. Lovrich, Washington State University
(Co-authors: Paul L. Shaffer, University of Arizona; Donald A. Yale, Northern Arizona University; Ronald A. Hopkins, Washington State University)

The Use of Candidate Self-Ratings as Validation Criteria
Bruce Davey, Connecticut State Personnel Division

Measuring Police Job Performance: A Comparison of Three Approaches
David P. Jones, Consulting Resources Corporation
(Co-authors: Mark Lifter, Arthur Young and Company; Peter Wentworth, Consulting Resources Corporation)

Applying Assessment Center Technology to Improving Performance Appraisal
Stephen Cohen, Assessment Designs, Inc.
(Co-author: Ward Thomas, Assessment Designs, Inc.)

The Utility of Similarity Judgments in Predicting Retranslation Errors
Michael A. McDaniel, Montgomery County Dept. of Personnel
(Co-author: Robert L. Holmgren, City of Chicago and Loyola University. Mr. Holmgren presented the paper.)

Quasi-Experimental Pilot Study of a PARTICIPATIVE Performance Appraisal System in Six Washington State Agencies: Effects Upon Job Satisfaction, Agency Climate, and Work Values

Dr. Lovrich's paper discussed a system of developing a uniform performance evaluation for the State of Washington. The first stage in this proje[illegible] was a survey of current employees and supervisors. This survey indicated that the majority of the individuals preferred a participative type of performance evaluation. The objective of the study was to set up a system for the State of Washington that was participative in nature. The study started out with a pilot project of three agencies that were brought into the participative mode and three agencies that were not. After a period of six months, these agencies were evaluated. Based on this evaluation, the system went statewide. After a period of 18 months using the statewide system, the researchers evaluated the success of the system, in three major areas:

1. Measurement of job climate
2. Measurement of job satisfaction
3. Measurement of work values

The general conclusions of the study indicated that there were significant improvements in the area of job climate and job satisfaction. However, there were no changes in the area of work values. Indications from both supervisors and employees indicate that they were more satisfied with the new participative system than with the old system.

The Use of Candidate Self-Ratings as Validation Criteria

Bruce Davey's paper dealt with the use of self-ratings or self-assessment as validation criteria. He indicated that some of the traditional problems of the criterion study included cost, small sample, restriction of range, and criterion development. Mr. Davey indicated that the use of self-ratings on the part of candidates would eliminate many of these problems and improve the data base. There are some problems with objectivity and halo effect but well-constructed rating forms can eliminate these. The Connecticut study dealt with results of their entry-level examination for their Professional Career Service. The rating form used by Connecticut included both factors directly related to the examination and some that were not. In general, the results indicated substantially significant correlations between the corresponding self-ratings and test parts. The results of this study indicate that self-ratings can be used as an effective device in evaluating test performance or applicant performance. This study also indicates that when care is used in developing self-ratings the results can be positive.

Measuring Police Job Performance: A Comparison of Three Approaches

Peter Wentworth discussed the results of a study using three different types of performance evaluations with 55 different metropolitan policy agencies. The three types of performance evaluations utilized for paired comparison, behavioral expectations, and graphic rating scales. All material was based on a critical incident job analysis. The major objective of the study was to see how the three rating systems interrelated. In general, the author felt that the paired comparison produced better results than behavioral expectations or the graphic rating scales. Reliability of the three systems was similar. The paired comparison, in general, tended to have more desirable traits. The study also looked at the differences between the supervisor and peer ratings, and generally found supervisory ratings were much more effective.

Applying Assessment Center Technology to Improving Performance Appraisal

Stephen Cohen presented a paper dealing with how assessment center technology can improve performance appraisal. Mr. Cohen urged to be aware that many of the techniques and procedures used in assessment centers can

have a valuable effect on performance evaluation. The important fact to remember is that for both assessment centers and for performance evaluation one observes behavior and that behavior must be relevant to the job in question and quantifiable. Thus technology and procedures utilized in an assessment center can indeed be transferred to performance evaluation.

The Utility of Similarity Judgments in Predicting Retranslation Errors

Robert Holmgren presented a paper on methods of retranslation. The objective of this is to eliminate redundant dimensions and to develop a rating scale which will maximize the amount of information that will be independent. Mr. Holmgren described some basic steps of development of behaviorally anchored rating scales. These include the identification dimensions, the writing of behavior examples, the retranslation and scaling of the behaviors, and the final development of the form. The concept of retranslation deals with the identification of the examples. The examples should be linked back to the dimensions that they were originally written to tap. If the statements are placed in the "wrong category" or dimension, the item has failed the retranslation process. Mr. Holmgren's contention was that unless one does an acceptable job of retranslation, the final results of the rating form would be inadequate.

Assessment Centers

Chair: Bill Owen, Department of State

Presentations:

Validity and Assessment Center Technology: One and the Same?
Stephen L. Cohen, Assessment Designs, Inc.

Assessment Centers: For Promotion Only?
Peter D. Wentworth, Wayne State University
(Co-author: James W. Thacker, Wayne State University)

New Directions in Government Assessment Center Applications
Frederick D. Frank, Assessment Designs, Inc.
(Co-author: Joe Sefcik, Assessment Designs, Inc.)

Effects of Assessor Training on Subsequent Performance in an Assessment Center
Frederick D. Frank, Assessment Designs, Inc.
(Co-author: Mike Struth, Assessment Designs, Inc.)

Validity and Assessment Center Technology: One and the Same?

This presentation began with a review of assessment center technology background. After discussing some of the reasons that assessment center technology and validity are <u>not</u> one and the same, some thoughts are presented

as to how validity can be improved. Content validity is presented as one of the underlying bases for assessment center success. Also discussed are the concepts of external and internal validity.

In conclusion, it was presented that the ingredients necessary to make validity and assessment center technology one and the same are:

(1) strict adherence to the theoretical rationale on which assessment center technology is based;
(2) comprehensive and legally defensible job analysis;
(3) content valid exercise simulations;
(4) appropriately relevant and sufficient assessor training;
(5) internally valid program administration;
(6) professional and ethical use of the results; and
(7) favorable return on investment.

Assessment Centers: For Promotion Only?

This presentation compared the utility of the assessment center process as a promotional device (traditional use) with its utility as an entry-level selection device.

The primary concerns in using an assessment center for entry-level selection focus on the time element and the cost factor. It was suggested that given the proper groundwork and circumstances, both of these problems could be successfully resolved.

The rest of the presentation traced the development and administration of an assessment center designed to select candidates for an entry-level sales position. The results obtained seemed to suggest that, at least in this instance, an assessment center worked well to select entry-level employees.

New Directions in Government Assessment Center Applications

This presentation discussed the importance of job analysis and its relationship to identifying the target position skills, the relationship of job analysis to instrument identification (and the need for innovative instruments), and, finally, the presentation discussed several nonmanagement applications and examples of types of instruments appropriate to those positions.

In summary, it was suggested that assessment centers can effectively be used for a variety of nonmanagerial applications. The key, however, is job analysis. With careful job analysis, the applications for assessment centers appear unlimited.

Effects of Assessor Training on Subsequent Performance

This presentation explored the crucial question of whether or not people who have been exposed to the assessment center process will subsequently have an advantage, in an actual assessment center, over those who have not been exposed to the process.

The subjects of the study designed to help answer this question were seventy-two managerial staff of the Federal Aviation Administration. The study was set up to determine whether intensive exposure (training as an assessor) to the assessment center process would influence subsequent assessment center performance.

The results of the study suggest that assessees who have been trained as assessors do not have any significant advantage over assessees who have not been trained as assessors.

SYMPOSIA

Clerical Selection and Cooperative Studies

Moderator: Stephen Berkley, Pennsylvania State Civil Service Commission

Participants: Charles Griffin, Pennsylvania State Civil Service Commission
Ronald Page, Control Data Corporation
Linda Hsu, University of Delaware
Theodore S. Darany, San Bernardino County Personnel Office

The clerical workforce is the largest occupational workforce category in the United States. Virtually every employer employs clerical workers, so it is extremely important that reliable and supportable selection devices are developed for this category of employee. This symposium presented and discussed the job analysis procedures, basic clerical dimensions defined for testing, tests developed, and validation procedures employed by the states of Minnesota, Delaware, and Pennsylvania.

Because of the large scope of these interjurisdictional studies, which included state and local governments, and their success and the similarity in findings, it is suggested that there may be high transferability possible for clerical tests.

Assessing Assessment Research

Moderator: Jacquel-Anne Chouinard, Commissioner of Personnel, State of Vermont

Participants: Glenn McClung, Denver Career Service Authority
Joel Wiesen, State of Massachusetts
Ken Millard, U.S. Office of Personnel Management

Discussant: Bruce Davey, Connecticut State Personnel Department

This symposium focused on identifying both common practices and deficiencies in the planning and conduct of assessment research and the steps which organizations can take to avoid or remedy such deficiencies.

Glenn McClung suggested that there are several readily identifiable problems with current research in the field of assessment which fall into two main categories: there is an overdemand for research and an undersupply of researchers who have both the technical and practical backgrounds for such study, and there is little effective direction and supervision of such activities because of little understanding by management of what research and researchers can do for an organization. Mr. McClung suggested that we can improve our research by taking some action on our own: adopt a multidisciplinary approach that exceeds the boundaries of only research design and statistics, learn and apply the work of others to our own work and

develop better communications with our own colleagues to learn what they want from research and how they are applying the work we give them.

Joel Wiesen addressed the evaluation of consultants' services and discussed how assessment methods can be transferred and applied to the evaluation of proposals for research. He described the development and application of dimensions and weights in a mini-study of interrater reliability of the evaluation of several research proposals. Mr. Wiesen concluded with some specific suggestions for inclusion in requests for proposals (RFPs) and contracts such as content requested by the RFP should be clearly related to evaluation areas so the proposals may be adequately evaluated, spell out personnel requirements, products expected should be clearly defined, reporting procedures and periodic approval by your organization and timetables.

Ken Millard discussed evaluating research products by both the assessment specialist and line management on the bases of technical soundness and apparent benefits. The need to address the question of transportability and extended application of our research to greater usefulness and other potential users was cited. Mr. Millard called for increased practical research and more publication of the results and application of our work.

PAPER SESSIONS

Innovations in Test Validation

Chair: Gerry Durovic, New York State Department of Civil Service

Presentations: Bootstrapping: An Elevation of Content Validity
Thomas A. Tyler, MEAS, Inc.

Probabilistic Selection: A Model for Continuous Validation
Leslie Melamed, Bell Canada
(Co-author: Paul D. Oelten, Bell Canada)

Process Evaluation of Public Welfare Worker Selection:
A Hypothesis Generating Technique
Peggy Goulding, Goulding, Martin and Associates

Bootstrapping: An Elevation of Content Validity

Thomas Tyler described a validation method he calls "bootstrapping" for use with job classes where the number of hires is too small for traditional empirical validation studies. The technique can be used to establish passing points and rankings on content validated selection procedures. It capitalizes on the ability of a regression equation to reduce individual errors in judgment and standardize the weights given to predictor variables. It is called bootstrapping because the relationship between the predictors (such as GRE or other test scores) and a criterion, which judges ratings based on the same predictors, is established directly from the applicant population and the regression equation is applied back to the same population to give candidates scores. The predicted scores generated by the regression equation have been found to be better predictors of future performance than the judges' direct ratings.

Probabilistic Selection: A Model for Continuous Validation

Leslie Melamed described a probabilistic selection model which dispenses with the traditional distinction between validation and test usage; where the validation phase is characterized by the need to ignore test results in the initial selection, and where the traditional test usage phase is characterized by a deterministic application of test results (involving a fixed cutoff). The model proposes that the selection of people be based either on their predicted criterion scores or on their probabilities of success on the job. Use of this model would allow one to use the test results from the first time the test is administered and continually update the validity of the test in use.

Process Evaluation of Public Welfare Worker Selection: A Hypothesis Generating Technique

Peggy Goulding described a hypothesis generating statistical technique which she used to evaluate the selection process for entry-level public welfare workers in a large state merit system. The technique is the Automatic Interaction Detector (AID) developed by the Institute for Social Research at the University of Michigan. She described the advantages of this technique over multiple linear regression, multiple discriminant function analysis, hierarchical grouping techniques.

She used the technique to profile the characteristics of candidates most likely to pass the written test, most likely to be certified for employment consideration and most likely to be hired. Through use of this technique, she identified possible conflicts in the steps of the selection procedure. Candidates eliminated at one step had characteristics that would have favored their retention later in the process.

Physical Testing

Chair: Sue Thrash, Wisconsin State Personnel Bureau

Presentations: The Development of Physical Ability Criterion Measures
David C. Myers, Advanced Research Resources Organization

First Encounters Acceptance Test (FEAT)--City of Tucson Firefighter Physical Abilities Test
Phil A. Carlin, City of Tucson Personnel Department

Development of the Physical Ability Test for the Classification Apprentice Fire and Rescue Officer
Donald A. Emmerich, City of Dallas Civil Service Department

Three presentations on physical testing were given in this paper session. Each presenter described the physical examination and the development of the examination.

The Development of Physical Ability Criterion Measures

A procedure for identifying the physical ability requirements of Army jobs was developed and used to collect data for a sample of incumbent soldiers drawn from four Military Occupational Specialties (MOSs). The findings indicated that the instruments and procedures designed to obtain perceived physical ability requirements were reliable. Several products emerged from the effort including profiles outlining the physical ability requirements of each MOS, a criterion task bank containing the more physically demanding tasks for each strength and stamina factor, and a procedures manual describing the steps necessary to establish physical ability requirements for additional MOSs. Research is currently underway to validate the procedure

so that the Army can begin to use the methodology to establish physical ability requirements for all MOSs on a judgmental rather than an empirical basis.

First Encounters Acceptance Test (FEAT)-- City of Tucson Firefighter Physical Abilities Test

Stressing work-task fidelity, FEAT is based on a job analysis which requested randomly selected firefighters to record tasks for a prior three-year period by type of fire and training drill. Each task was then rated on a Scale of Importance ranging from Routine to Life Saving by the Fire Department's Testing Committee.

Test activities selected were a hose roll, hose advance, ladder set-up, attic crawl, sledge hammer event, stairs event, and dummy drag.

FEAT was pretested using 34 randomly selected Firefighters and 16 civilians including 6 women. A standard was established by combining findings from both samples. Three firefighter personnel and eight civilians failed to complete the test. None of the successful eight civilians were women. Those successfully completing FEAT were most often cardiovascular development enthusiasts with regular jogging, bicycling, or swimming being the most popular activities.

FEAT was administered to 255 candidates over a five-day period. The average day required 9½ hours. A Paramedic unit was on call and was required at least once a day. Six monitors were assigned to help applicants put on equipment, provide verbal instruction, and assure safe conditions.

Results

Candidates

	Attempting FEAT	Completing FEAT	Percentage Successful
Men	239	194	81
Women	16	4	25
TOTAL	255	198	78

Average Time: 8 minutes 38 seconds

Development of the Physical Ability Test for the Classification Apprentice Fire and Rescue Officer

The City of Dallas conducted a job analysis to determine tasks most important for Apprentice Fire and Rescue Officer. Tasks selected for inclusion on the test were those most able to be simulated in a testing environment, not requiring use of specialized equipment or training, and not presenting a serious safety hazard to applicants.

The test was developed, pretested on incumbents, and refined as new information was acquired. A study of personal characteristics, such as age, height, weight, race, and test performance was conducted. Significant differences were found in the personal characteristics of whites, blacks, and Mexican-Americans and in their performances on two test parts and on total time.

The results for the tests suggested that height and weight screening may have substantial predictive utility for white and black candidates, but not for Mexican-Americans, whose height had no apparent relationship to test performance.

Issues in Test Fairness and Affirmative Action

Chair: Lance Seberhagen, Seberhagen and Associates, Inc.

Presentations: The Impact of Recent Court Decisions on Personnel Selection and Affirmative Action
Jeanette J. Lim, Development Research Corporation

The Real Impact of Affirmative Action Programs
Neuman F. Pollack, Center for the Study of Administration, NOVA University, Miami, Florida

The Impact of Recent Court Decisions on Personnel Selection and Affirmative Action

Jeanette Lim summarized several major court decisions in an effort to identify the positions and remedies of various courts concerning affirmative action.

Regarding permissibility of affirmative action (AA), some findings were:

Bakke: AA is permissible if there has been prior adverse impact and if there is a legitimate purpose.

Weber: Nothing requires preferential treatment, but nothing keeps the court from permitting it.

This paper session concentrated on describing alternative methods of selection. Gary Stormoem and Amiel Sharon discussed the development of ratings of past performance as a method for selection.

Cynthia Amick described her state's use of trainee classifications as an alternate method of appointing individuals to permanent Civil Service classifications.

Proficiency examinations for Clinical Laboratory Technicians were described by Maureen Kaley. Their study compared a written test format with a practical test format.

Other cases: The court ruled that Congress was able to determine when quotas were required for the general welfare and to enforce through legislation. The limited use of race and ethnicity was a legitimate use of congressional prerogatives to remedy discrimination.

Regarding the remedies that various courts have proposed, some findings were:

Castro v. Beecher: Use of priority pools and preferential hiring.

Bridgeport F.D.: Lower cutoff score to eliminate adverse impact and hire randomly to avoid adverse impact of ranking.

Pennsylvania State Troopers: Temporary hiring goal requiring one minority for every two whites until 9 percent representation.

Mississippi Highway Safety Patrol: Every black who passed the test was ordered to hired.

Toledo Police Department: Seniority use reduced from a five and two year requirement for promotions to one year. Court ordered validation, but appeal of seniority reduction remanded this to district court.

Los Angeles P.D.: Court enjoined LEAA from denying or discontinuing grants when the city refused to set up affirmative action goals and timetables.

The Real Impact of Affirmative Action Programs

This presentation was a discussion of the experiences of the conflicts of racial tension and affirmative action programs in the state of Florida.

Alternative Selection Techniques

Chair: Bruce Davey, Connecticut State Personnel Department

Presentations: Task Performance Self-Evaluations: An Alternative Selection Procedure to Traditional Experience and Training Ratings
Barry Farrell, Minnesota Department of Personnel
(This paper was presented by Gary K. Stormoem, Minnesota Department of Personnel)

Alternative Selection Method for Disadvantaged Applicants
Cynthia L. Amick, Pennsylvania State Civil Service Commission

A Methodology for the Development of Practical Examination and Some Statistical Findings
Maureen M. Kaley, Professional Examination Service

Development of Alternative Selection Procedures for Federal Administrative Judge
Amiel Sharon, U.S. Office of Personnel Management

Discussant: Charles Schultz, Washington State Department of Personnel

PRESIDENTIAL ADDRESS

Automating the Examination Process in the 80s

Theodore S. Darany
San Bernardino County Personnel Division

In the first part of this presentation, outgoing IPMAAC President Darany offered a mind-boggling perspective on the growth of the computer field over the past 25 years. Some examples of the stunning parallels drawn are:

- If a computer using the vacuum tube technology of 25 years ago were built to match the computing power of a modern hand-held SR-59 calculator, it would have to be as large as New York City.
- If the efficiency and expense of cars had improved at the same rate as computers, a Rolls Royce today would cost about $3 (instead of $80,000) and get 3,000,000 miles to a gallon (instead of 12).
- If the aerospace field had progressed at the same rate as computer technology, then the time distance between the Wright brothers and the moon landing would have been 4½ years.

The rapidity of change in this field heralds further rapid advances. Dr. Darany discussed what this means for the examination process today and speculated on what changes might occur in the near future, in his home jurisdiction of San Bernardino and nationwide. Some of the trends discussed were:

Computer-Assisted Job Analysis has brought the classification and examination sections together in cooperative studies in San Bernardino County. Using the CODAP system, they studied 2,400 clerical positions in nine months, producing data useful for many personnel functions.

Computer-Assisted Examining includes not only computerized item banks which store items, update item usage and quality statistics and print exams, but also applicant tracking from initial application all the way through to mailing results to candidates, certification, and EEO reporting.

Computer Test Proctoring in which the computer actually administers the examination to the candidate via terminal (or even by a "talking" computer). Shorter tests tailored to each candidate's ability level can become more commonplace when remote terminals become as common as telephones.

INVITED ADDRESS
Sponsored by the Great Lakes Assessment Council

Host: Frank Smith, Department of Personnel, City of Detroit

Personnel Assessment from the Perspective of Work Adjustment

Rene V. Dawis, University of Minnesota

Rene Dawis discussed personnel assessment from the perspective of his theory of work adjustment. According to Dawis, the individual's perception of the work environment includes the interpersonal, the organizational, the social, as well as the physical. The work organization's perception of the individual is in terms of the tasks that the individual can perform in order to achieve organizational objectives. Professor Dawis suggested that it is important to match the individual to the work environment and that the degree of correspondence between the individual's skills and work environment is reflected by satisfaction.

In theoretical terms, the Theory of Work Adjustment predicts degree of individual and organizational satisfaction from the match between individual needs and skills and organizational task requirements and reinforcers.

According to Professor Dawis, personnel assessment is best for screening out unqualified candidates rather than ranking or grouping. He also indicated that the capability to perform the job is what should be assessed, and content validity is a necessary condition for the assessment procedure.

PAPER SESSION

Various Approaches to Measuring the Cost-Effectiveness of Tests

Chair: David Friedland, City of Los Angeles

Presentation: Three Methods of Measuring Cost Effectiveness
Charles B. Schultz, Washington State Department of Personnel

Discussant: David Friedland, City of Los Angeles

Three Methods of Measuring Cost Effectiveness

Two measures of cost effectiveness showed the benefit of selection by test when compared to chance was valued at about $3,000 per year per person hired. The third method showed a value of $64,000 per year per person; however, this is the difference in money collected rather than difference in value of the work. The criterion was the amount of child support money collected by Support Enforcement Officers.

The latter method, called the cost accounting method, overestimated the value of the work performed. The other methods were referred to as shortcut methods. The author's method assumed that the value of the best performer was three times that of the worst performer in a work group and that the average performer was worth the prevailing rate. The data showed that the best performers collected more than three times as much support money than the poorest.

The second shortcut method was one designed by Frank Schmidt. He had supervisors estimate the value of workers' performance that was one standard deviation above and below the mean. The two shortcut methods yielded similar results and gave more realistic indications of the value of work performed than the cost accounting method.

Two validity studies had shown the test that was used to be valid. However, the validity coefficient for the employees for whom data were available failed to reach significance. The author advised against relying on employer-collected data, since this employer collected data selectively on the poorer employees.

SYMPOSIA

Assessing Writing Skills for Employment Selection

Moderator: Barbara Showers, State of Wisconsin, Division of Personnel

Participants: Joe M. Steele, American College Testing Program
Jack R. Lawton, State of Wisconsin, Division of Personnel

Discussant: Ronald Ash, University of South Florida

Joe Steele discussed the American College Testing Program's College Outcome Measures Project (COMP) and his findings regarding the assessment of writing skills for that project. Writing assessment procedures developed include a test composed of three different role-playing writing tasks and a qualitative rating procedure which includes judgments of audience recognition, organization, and language usage. The measure was found to be a reliable, content valid assessment of writing performance of college graduates. Longitudinal studies are in progress to determine if COMP is predictive of effective functioning in various adult roles.

Jack Lawton discussed the development and reliability studies of written communication rating scales for use in personnel selection. He found that the most reliable ratings were obtained with a three-scale analytical assessment of multiple writing samples similar to that proposed for Steele's project. He found that it was not necessary to have raters with special expertise in written communication as long as they were trained in the use of the scales. He concluded that reliable assessment of written communication skills could be achieved for personnel selection tests and would be a cost-effective and content valid choice for jobs with somewhat limited applicant numbers and a documented need for written communication skills.

Ron Ash discussed the two studies and the issue of measuring written communication. He concluded that both studies showed that the typical problem of low reliability for essay tests can be overcome. The "formula for success" includes: using tasks of optimal length, high standardization of tasks, use of evaluators with content expertise, high quality scale development, rater training, and grading techniques to minimize bias, such as disregard penmanship, rate one question at a time, reorder responses periodically.

Alternative Selection Procedures for Entry-Level Positions

Moderator: Robert L. Hannan, U.S. Office of Personnel Management

Participants: Robert L. Hannan, U.S. Office of Personnel Management
Ronald L. Karren, U.S. Office of Personnel Management
Anthony J. Mento, U.S. Office of Personnel Management
Charles N. MacLane, U.S. Office of Personnel Management

Discussant: Sherry Turpenoff, U.S. Office of Personnel Management

The field of personnel selection is currently focusing its attention on ways to reduce adverse impact. Since written tests have been most widely used, alternative testing techniques are being explored. This symposium discussed a number of such methods. Emphasis was given to the problems and possibilities of using these methods for selection to entry-level white collar jobs.

Mr. Hannan discussed reasons why selection methods based solely on work skills or solely on individual or situation-based work motivation theories have not been able to predict job success with low adverse impact. He proposed a framework for developing and implementing a person-job environment congruence model of expected work motivation as a selection device for entry-level positions. The method is based on computer-matching of candidate work motivation profiles to job environment work motivation profiles. The method requires new job analysis techniques, but the selection device could be given as a computer-scored test.

Mr. Karren proposed a utility model for identifying and combining selection devices which takes adverse impact into account by assigning a negative dollar value to represent social cost. The model would enable the selection specialist to compare and choose selection devices in order to maximize validity and utility and minimize adverse impact.

Mr. Mento discussed the unique qualities of the interview, the assessment center, and the work sample as techniques for entry-level selection with low adverse impact. Although the assessment center and work sample are often viewed as too costly, Mr. Mento pointed out that expenses may be minimized in the long run if the quality of the applicant is substantially increased, resulting in increased competent productivity and reduced turnover. The interview is more popular, but is more prone to unreliability and bias unless carefully constructed.

Mr. MacLane described a study of the use of a form of work sample test known as the miniature training and evaluation examination. The examination was constructed to select individuals into upward-mobility positions--technician jobs that could lead to professional opportunities. The jobs were designed to reduce adverse impact by not requiring formal training prior to hire. The work samples were self-contained units similar to the training and interpersonal situations which candidates would encounter on the job. The test was found to be adequately reliable and empirically valid, but was not completely successful in eliminating adverse impact on blacks and females. It was hypothesized that degree of education was still affecting test performance between groups.